

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІСЬКОГО ГОСПОДАРСТВА імені О. М. БЕКЕТОВА

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
з дисципліни
«Моделювання та прогнозування стану навколишнього середовища»

до виконання контрольної роботи

РОЗРОБКА РЕГРЕСІЙНОЇ МОДЕЛІ ВПЛИВУ СКИДУ СТИЧНИХ ВОД
ПІДПРИЄМСТВА НА ЯКІСТЬ РІЧКОВОЇ ВОДИ В КОНТРОЛЬНОМУ
СТВОРІ

(для студентів 6 курсу заочної форми навчання спеціальності
263 – Цивільна безпека)

Методичні рекомендації з дисципліни «Моделювання та прогнозування стану навколишнього середовища» до виконання контрольної роботи «Розробка регресійної моделі впливу скиду стічних вод підприємства на якість річкової води в контрольному створі» (для студентів 6 курсу заочної форми навчання спеціальності 263 – Цивільна безпека) / Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова ; уклад. В. О. Бараннік. – Харків : ХНУМГ ім. О. М. Бекетова, 2019. – 22 с.

Укладач канд. фіз.-мат. наук, доц. В. О. Бараннік

Рецензент канд. техн. наук, доц. Є. Г. Пономаренко

*Рекомендовано кафедрою інженерної екології міст,
протокол № 1 від 29 серпня 2019 р.*

ЗМІСТ

Вступ.....	4
1 Загальні положення.....	5
2 Послідовність простого регресійного аналізу.....	6
2.1 Візуалізація розподілу даних вимірювань.....	6
2.2 Кореляційний аналіз.....	7
2.3 Розрахунок параметрів регресії.....	8
2.4 Визначення властивостей шуму і похибок оцінок параметрів.....	9
2.5 Прогнози за лінійною парною регресією.....	11
3 Приклад простого регресійного аналізу.....	11
3.1 Візуалізація розподілу даних вимірювань.....	12
3.2 Кореляційний аналіз.....	13
3.3 Розрахунок параметрів регресії.....	14
3.4 Визначення властивостей шуму і похибок оцінок параметрів.....	14
3.5 Прогнози за лінійною парною регресією.....	16
3.6 Висновки.....	17
4 Вимоги до виконання роботи.....	18
Список рекомендованої літератури.....	19
Додатки.....	20

ВСТУП

Метою контрольної роботи є поглиблення теоретичних знань та надбання практичних навичок у розрахунках параметрів простих регресійних моделей процесів навколишнього середовища за даними синхронних спостережень та графічному відображенні (візуалізації) даних спостережень і розрахунків.

Об'єктом розгляду є ділянка ріки проміж випуском стічної води і контрольним створом (КС), де відбуваються процеси розбавлення і трансформації забруднюючих речовин (домішок), що надходять зі стічною водою скиду та від вище розташованих джерел забруднення.

У завданні на контрольну роботу наводяться дані синхронних вимірювань концентрації домішки на випуску стічної води і в КС річки на відстані 500 м від місця випуску.

Результатом роботи мають бути розраховані:

- оцінки параметрів парної лінійної регресійної моделі залежності вмісту домішки в КС річки від її вмісту в стічній воді;
- залишки регресії;
- характеристики «шуму»;
- прогнозні величини концентрації домішки в КС за наданою її концентрацією в стічній воді на випуску у річку.

На підставі проведених розрахунків мають бути визначені умови існування парної лінійної регресії даних вимірювань та параметри регресії.

Пояснювальна записка до контрольної роботи має включати зміст, п'ять розділів, три вказаних графіка регресійного аналізу даних на двох рисунках, дві таблиці з вихідними даними й результатами розрахунків.

1 ЗАГАЛЬНІ ПОЛОЖЕННЯ

Скиди виробничих, побутових і сільськогосподарських стічних вод у поверхневі води вносять чуттєві зміни в їх гідрохімічний і біологічний режим, змінюючи якість води, порушуючи життєдіяльність рослинних і тваринних організмів. Тому витрати скиду стічних вод і вміст забруднюючих речовин (домішок) у стічних водах підлягають нормуванню залежно від асимілюючої здатності поверхневих вод. Нормативи гранично допустимих скидів розраховуються і встановлюються з використанням математичних моделей впливу скидів на якість поверхневих вод для гарантованих з певною забезпеченістю умов скидання. Найбільш поширеною моделлю розрахунку впливу окремого скиду стічної води на якість води річки є модель матеріального балансу забруднюючої речовини у вигляді

$$C_{\kappa\sigma} = \frac{C_{\sigma\tau} - C_{\phi}}{n} + C_{\phi} . \quad (1)$$

де $C_{\kappa\sigma}$ – вміст домішки в КС;

$C_{\sigma\tau}$ – вміст домішки в стічній воді скиду;

C_{ϕ} – фоновий вміст домішки в річці безпосередньо перед місцем розташування випуску стічної води;

n – кратність розбавлення стічної води річковою на шляху від місця випуску до КС.

Проте оцінка впливу скиду на вміст забруднюючої речовини у контрольному створі за моделлю (1) є досить складною проблемою через те, що, наприклад, кратність розбавлення n є величиною, складним чином залежною від місцевих змінних гідродинамічних умов переносу речовини річкою. Альтернативою є побудова статистичної моделі оцінки впливу скиду, якщо існують дані регулярних синхронних гідрохімічних спостережень (вимірювань) вмісту домішки в стічній воді та в КС, що мають проводитися на випусках з великими річними об'ємами скиду стічної води. Саме модель (1) дозволяє висловити припущення (сформулювати гіпотезу) щодо структури

статистичного зв'язку $C_{\kappa c}$ і C_{cm} :

$$C_{\kappa c} = n^{-1}C_{cm} + (1 - n^{-1})C_{\phi} = \theta_0 + \theta_1 C_{cm} + e = f(C_{cm}) + e,$$

де $f(C_{cm}) = \theta_0 + \theta_1 C_{cm}$ є лінійна регресія величини $C_{\kappa c}$ за величиною C_{cm} ;

$\theta_1 = n^{-1}$ і $\theta_0 = (1 - n^{-1})C_{\phi}$ є параметрами регресії;

e – внесок випадкових факторів («шум»), що визначають неконтрольований вплив фонові концентрації, умов розбавлення стічної води у річці, похибок вимірювань концентрації домішки тощо.

Задачами регресійного аналізу (розробки регресійної моделі) є:

а) перевірка припущення, що регресія має вигляд

$$C_{\kappa c} = \theta_0 + \theta_1 C_{cm}; \quad (2)$$

б) визначення величин регресійних параметрів θ_0 і θ_1 , а також

в) визначення характеристик «шуму» e та його впливу на точність визначення параметрів регресії.

Вихідними даними для регресійного аналізу має слугувати репрезентативна вибірка одночасних вимірювань

$$(C_{cm,1}, C_{\kappa c,1}), (C_{cm,2}, C_{\kappa c,2}), \dots, (C_{cm,N}, C_{\kappa c,N})$$

концентрації $C_{cm,i}$ забруднюючої речовини у стічній воді на випуску у річку і концентрації $C_{\kappa c,i}$ забруднюючої речовини у контрольному створі нижче за течією (N – розмір вибірки). Такі вибірки можуть бути утворені, наприклад, за планом простого випадкового відбору або за планом систематичного випадкового відбору (серійний відбір).

2 ПОСЛІДОВНІСТЬ ПРОСТОГО РЕГРЕСІЙНОГО АНАЛІЗУ

2.1 Візуалізація розподілу даних вимірювань

Спочатку треба візуально упевнитися, що розподіл даних вимірювань має ознаки лінійної регресії. Це досягається побудовою точкового графіка розподілу результатів парних вимірювань $(C_{cm,i}; C_{\kappa c,i})$ у прямокутній системі координат, де вздовж вісі абсцис відкладають величини $C_{cm,i}$, а вздовж осі

ординат – величини $C_{kc,i}$. Якщо розподіл точок утворює видовжену вздовж певного напрямку «хмаринку» з нахилом до осей координат, то сподівання на наявність лінійної регресії (1) мають укріпитися. Якщо точки утворюють хаотичний клубок без певного нахилу «хмаринки», то сподівання на наявність регресії мають послаблюватись.

2.2 Кореляційний аналіз

За виявленими візуальними ознаками наявності лінійної регресії рекомендується виконати простий кореляційний аналіз даних для підтвердження цієї наявності. Для цього спочатку треба розрахувати величини кількох статистик (тобто величин, що визначаються виключно через дані вимірювань), а саме:

– вибіркове середнє величин $C_{cm,i}$

$$\overline{C}_{cm} = \frac{1}{N} \sum_{i=1}^N C_{cm,i} ;$$

– вибіркове середнє величин $C_{kc,i}$

$$\overline{C}_{kc} = \frac{1}{N} \sum_{i=1}^N C_{kc,i} ;$$

– вибіркове середнє добутків $C_{cm,i} C_{kc,i}$

$$\overline{C_{cm} C_{kc}} = \frac{1}{N} \sum_{i=1}^N C_{cm,i} C_{kc,i} ;$$

– середнє квадратичне (стандартне) відхилення величин $C_{cm,i}$

$$s_{cm} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (C_{cm,i} - \overline{C}_{cm})^2} ;$$

– середнє квадратичне (стандартне) відхилення величин $C_{kc,i}$

$$s_{kc} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (C_{kc,i} - \overline{C}_{kc})^2} ;$$

– вибірковий коефіцієнт кореляції даних вимірювань

$$R = \frac{N}{N-1} \cdot \frac{\overline{C_{cm} C_{kc}} - \overline{C}_{cm} \overline{C}_{kc}}{s_{cm} s_{kc}} .$$

Далі перевіряють «нульову гіпотезу», тобто припущення, що коефіцієнт кореляції дорівнює нулю і регресія (1) відсутня. Перевірку нульової гіпотези проводять порівнюючи критерій достовірності T з критерієм Стюдента t_s з $\nu = N - 2$ ступенями вільності і рівнем значущості $\alpha = 0,05$ (див. Додаток А). Величину критерію достовірності розраховують за формулою

$$T = |R| \sqrt{\frac{N-2}{1-R^2}}.$$

Якщо $T > t_s$, то нульова гіпотеза відкидається, що дозволяє констатувати наявність кореляційного зв'язку, а значить і регресії (2), між вимірюваннями. Також, визначається величина коефіцієнта детермінації для регресії, що дорівнює R^2 . За таким висновком є сенс продовжити розробку вказаної регресійної моделі.

2.3 Розрахунок параметрів регресії

Регресійний аналіз даних спостережень виконують за положенням, що найліпшою оцінкою $\hat{\theta}_0$, $\hat{\theta}_1$ невідомих величин параметрів θ_0 , θ_1 , регресії (2) є така, що надає мінімального значення сумі квадратів залишків регресії (*метод найменших квадратів*):

$$ЗСВ = \sum_{i=1}^N e_i^2 = \sum_{i=1}^N (C_{\kappa\epsilon,i} - f(C_{\epsilon m,i}))^2 = \sum_{i=1}^N (C_{\kappa\epsilon,i} - \hat{\theta}_0 - \hat{\theta}_1 C_{\epsilon m,i})^2 = \min,$$

де

$$e_i = C_{\kappa\epsilon,i} - \hat{\theta}_0 - \hat{\theta}_1 C_{\epsilon m,i} \quad (3)$$

є залишок регресії для i -го вимірювання концентрацій речовини.

Математичний аналіз задачі мінімізації ЗСВ дає наступні формули для визначення *точкових оцінок* (одним числом) параметрів регресії:

$$\hat{\theta}_1 = \frac{S_{\kappa\epsilon}}{S_{\epsilon m}} \cdot R; \quad (4)$$

$$\hat{\theta}_0 = \bar{C}_{\kappa\epsilon} - \hat{\theta}_1 \cdot \bar{C}_{\epsilon m}. \quad (5)$$

З урахуванням формул (4, 5) рівняння регресії (2) може бути надано у наступному вигляді:

$$\frac{C_{kc} - \bar{C}_{kc}}{S_{kc}} = R \cdot \frac{C_{cm} - \bar{C}_{cm}}{S_{cm}}.$$

Оцінку якості рівняння регресії в цілому виконують за допомогою F -критерію Фішера:

$$F = \frac{R^2}{1 - R^2} (N - 2).$$

Це число порівнюють з табличним значенням $F_{\text{табл}}$, що відповідає наступним параметрам $k_1 = 1$, $k_2 = N - 2$, $\alpha = 0,05$ (дод. А). Якщо $F > F_{\text{табл}}$, то визнають статистичну значимість рівняння регресії в цілому.

Очевидно, що графіком рівняння $C_{kc} = f(C_{cm}) = \hat{\theta}_0 + \hat{\theta}_1 C_{cm}$ в прямокутній системі координат $C_{cm} OC_{kc}$ є пряма з нахилом $\hat{\theta}_1$, яка проходить через точку $(\bar{C}_{cm}, \bar{C}_{kc})$.

2.4 Визначення властивостей шуму і похибок оцінок параметрів

Після отримання оцінок параметрів регресії розраховують залишки регресії за формулою (3) у кількості N одиниць. і визначають їх середнє квадратичне (стандартне) відхилення за формулою

$$s_e = \sqrt{\frac{1}{N - 2} \sum_{i=1}^N e_i^2}, \quad (6)$$

що є важливою статистичною характеристикою інтенсивності шуму.

Якщо залишки регресії є дійсно наслідком впливу дії випадкових факторів на результати вимірювань (а саме таке припущення є основою застосування наведених критеріїв оцінки статистичної значущості статистичних параметрів), то, по-перше, розташування їх графічних образів (точок) у прямокутній системі координат з осями (iOe_i) не повинно мати ознак будь-якої упорядкованості (точки мають бути хаотично розміщеними) і, по-друге, точки, які виходять далеко за межі горизонтальної полоси навколо вісі абсцис з границями $e = t_s s_e$ і $e = -t_s s_e$ мають вважатися такими, що порушують однорідність вибірки спостережень, і підлягають видаленню з вибірки.

Величина (6) також визначає похибки оцінок статистичних параметрів.

Так, стандартна похибка визначення параметра регресії θ_0 розраховується за формулами

$$m(\theta_0) = s_e \cdot \frac{\sqrt{\overline{C_{cm}^2}}}{s_{cm} \sqrt{N-1}};$$

$$\overline{C_{cm}^2} = \frac{1}{N} \sum_{i=1}^N C_{cm,i}^2.$$

За схожою формулою розраховується стандартна похибка визначення параметра регресії θ_1 :

$$m(\theta_1) = s_e \cdot \frac{1}{s_{cm} \sqrt{N-1}}.$$

Статистичну значущість оцінок параметрів регресії визначають звичайним чином, перевіряючи перевищення величин $\hat{\theta}_0 / m(\theta_0)$ і $\hat{\theta}_1 / m(\theta_1)$ табличної величини критерію Стюдента t_s з $\nu = N - 2$ ступенями вільності і рівнем значущості $\alpha = 0,05$ (дод. Б). Відповідно, *інтервальні оцінки* справжніх величин параметрів регресії визначають довірчими інтервалами так

$$\hat{\theta}_0 - t_s m(\theta_0) \leq \theta_0 \leq \hat{\theta}_0 + t_s m(\theta_0); \quad (7)$$

$$\hat{\theta}_1 - t_s m(\theta_1) \leq \theta_1 \leq \hat{\theta}_1 + t_s m(\theta_1), \quad (8)$$

або так

$$\theta_0 = \hat{\theta}_0 \pm t_s m(\theta_0);$$

$$\theta_1 = \hat{\theta}_1 \pm t_s m(\theta_1),$$

або так

$$\theta_0 \in (\hat{\theta}_0 - t_s m(\theta_0); \hat{\theta}_0 + t_s m(\theta_0));$$

$$\theta_1 \in (\hat{\theta}_1 - t_s m(\theta_1); \hat{\theta}_1 + t_s m(\theta_1)).$$

Середня похибка апроксимації даних вимірювань рівнянням регресії (у відсотках) визначається за формулами

$$A = \frac{1}{N} \sum_{i=1}^N \left| \frac{C_{kc,i} - \hat{f}_r(C_{cm,i})}{C_{kc,i}} \right| \cdot 100 \%;$$

$$\hat{f}_r(C_{cm,i}) = \hat{\theta}_0 + \hat{\theta}_1 C_{cm,i}.$$

2.5 Прогноз за лінійною парною регресією

Розроблену регресійну модель можна використовувати для прогнозу величини концентрації домішки, що можна очікувати в КС річки, якщо її концентрація в стічній воді на скиді буде дорівнювати певній величині $C_{cm,n}$. У цьому випадку розрізняють два види прогнозів:

а) прогноз середньої (за даними вимірювань) концентрації $\bar{C}_{кс,n}$ домішки в КС, якщо її концентрація в стічній воді дорівнює $C_{cm,n}$ і

б) прогноз результату вимірювання концентрації $C_{кс,n}$ в КС, коли її концентрація в стічній воді дорівнює $C_{cm,n}$.

Точкові оцінки для обох прогнозів надаються однаковою формулою

$$\bar{C}_{кс,n} = C_{кс,n} = \hat{\theta}_0 + \hat{\theta}_1 C_{cm,n}.$$

Проте інтервальні оцінки прогнозованих величин дещо розрізняються і надаються відповідними довірчими інтервалами, а саме:

$$\bar{C}_{кс,n} = \hat{\theta}_0 + \hat{\theta}_1 C_{cm,n} \pm t_s s_e \sqrt{\frac{1}{N} + \frac{(C_{cm,n} - \bar{C}_{cm})^2}{(N-1)s_{cm}^2}};$$

$$C_{кс,n} = \hat{\theta}_0 + \hat{\theta}_1 C_{cm,n} \pm t_s s_e \sqrt{1 + \frac{1}{N} + \frac{(C_{cm,n} - \bar{C}_{cm})^2}{(N-1)s_{cm}^2}}.$$

3 ПРИКЛАД ПРОСТОГО РЕГРЕСІЙНОГО АНАЛІЗУ

Розглянемо приклад виконання регресійного аналізу за наступним завданням.

Варіант Х.

За даними вимірювань концентрацій домішки у стічній воді та в контрольному створі річки (табл. 1):

а) розробити лінійну модель парної регресії $C_{kc} = \theta_0 + \theta_1 C_{cm}$;

б) надати прогноз концентрації домішки в контрольному створі, якщо концентрація домішки на випуску стічної води буде дорівнювати $C_{cm,n} = 180 \text{ мг/дм}^3$.

Таблиця 1 – Дані вимірювань концентрації домішки (мг/дм³)

i	$C_{cm,i}$	$C_{kc,i}$	i	$C_{cm,i}$	$C_{kc,i}$	i	$C_{cm,i}$	$C_{kc,i}$
1	231	38,5	11	249	38,4	21	290	38,5
2	366	39,9	12	283	39,2	22	297	38,9
3	255	39,1	13	374	40,2	23	404	40,1
4	237	38,7	14	326	39,0	24	340	39,1
5	268	38,6	15	308	39,6	25	304	39,1
6	347	39,8	16	301	38,8	26	250	38,2
7	386	39,8	17	265	39,2	27	300	39,6
8	366	39,5	18	375	40,0	28	328	39,2
9	276	39,0	19	352	39,5	29	251	38,5
10	235	38,2	20	369	40,2	30	292	38,7

3.1 Візуалізація розподілу даних вимірювань

Рисунок 1 дає уявлення щодо розташування точок, що представляють дані парних вимірювань концентрації домішки, у прямокутній системі координат. Очевидно, що наведені на рисунку точки можна сприймати розміщеними щільно в межах уявного еліпсу, у якого більша вісь має певний нахил відносно горизонтальної вісі. Ця обставина дозволяє розраховувати на те, що між змінними C_{kc} і C_{cm} існує лінійна статистична залежність у вигляді регресії.

3.2 Кореляційний аналіз

Для виконання кореляційного аналізу розраховуємо кілька простих статистик:

– вибіркове середнє величин $C_{cm,i}$

$$\bar{C}_{cm} = \frac{1}{N} \sum_{i=1}^N C_{cm,i} = 307,5 \text{ мг/дм}^3;$$

– вибіркове середнє величин $C_{kc,i}$

$$\bar{C}_{\kappa\kappa} = \frac{1}{N} \sum_{i=1}^N C_{\kappa\kappa,i} = 39,17 \text{ мг/дм}^3;$$

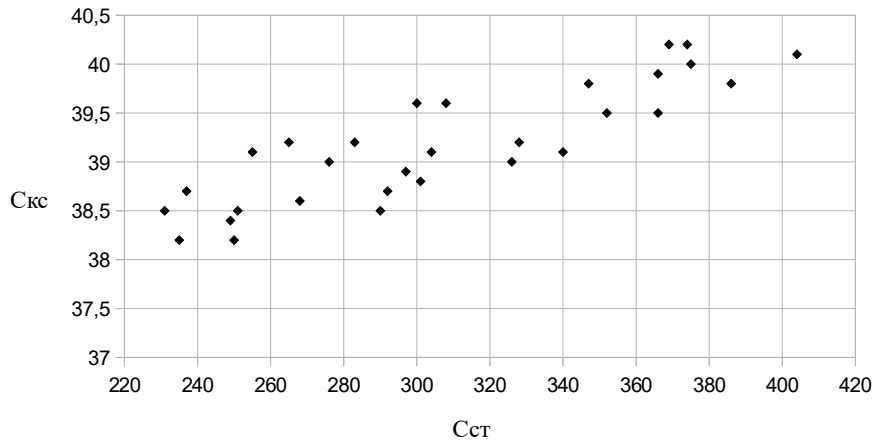


Рисунок 1 – Результати 30 вимірювань концентрацій домішки у стічній воді ($C_{cm,i}$; мг/дм³) і в контрольному створі річки ($C_{\kappa\kappa,i}$; мг/дм³)

– вибіркове середнє добутків $C_{cm,i} C_{\kappa\kappa,i}$

$$\overline{C_{cm} C_{\kappa\kappa}} = \frac{1}{N} \sum_{i=1}^N C_{cm,i} C_{\kappa\kappa,i} = 12\,070 \text{ мг}^2/\text{дм}^6;$$

– середнє квадратичне (стандартне) відхилення величин $C_{cm,i}$

$$s_{cm} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (C_{cm,i} - \bar{C}_{cm})^2} = 50,83 \text{ мг/дм}^3;$$

– середнє квадратичне (стандартне) відхилення величин $C_{\kappa\kappa,i}$

$$s_{\kappa\kappa} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (C_{\kappa\kappa,i} - \bar{C}_{\kappa\kappa})^2} = 0,600\,7 \text{ мг/дм}^3;$$

– вибірковий коефіцієнт кореляції даних вимірювань

$$R = \frac{N}{N-1} \cdot \frac{\overline{C_{cm} C_{\kappa\kappa}} - \bar{C}_{cm} \bar{C}_{\kappa\kappa}}{s_{cm} s_{\kappa\kappa}} = 0,867\,7.$$

Далі розраховуємо величину критерію достовірності

$$T = |R| \sqrt{\frac{N-2}{1-R^2}} = 9,237.$$

За таблицею (дод. А) знаходимо табличну величину критерію Стьюдента $t_s = 2,048\,4$ з $\nu = N - 2 = 28$ ступенями вільності і рівнем значущості $\alpha = 0,05$. У

зв'язку з тим, що $T > t_s$, то нульова гіпотеза відкидається, так що приймаємо наявність кореляційного зв'язку, а значить і регресії (2) між змінними C_{kc} і C_{cm} . Також, визначаємо величину коефіцієнта детермінації

$$R^2 = 0,753.$$

3.3 Розрахунок параметрів регресії

Розраховуємо точкові оцінки параметрів регресії:

$$\hat{\theta}_1 = \frac{S_{kc}}{S_{cm}} \cdot R = 0,010\ 25;$$

$$\hat{\theta}_0 = \bar{C}_{kc} - \hat{\theta}_1 \cdot \bar{C}_{cm} = 36,02 \text{ мг/дм}^3.$$

Таким чином, рівняння регресії має вигляд

$$C_{kc} = 0,010\ 25 \cdot C_{cm} + 36,02. \quad (7)$$

Для перевірки статистичної значущості рівняння регресії розраховуємо величину F -критерія Фішера:

$$F = \frac{R^2}{1 - R^2} (N - 2) = 85,32.$$

Ця величина значно перевищує табличне значення $F_{\text{табл}} = 4,20$, яке відповідає наступним параметрам $k_1 = 1$, $k_2 = N - 2 = 28$, $\alpha = 0,05$ (дод. А), що визначає статистичну дійсність лінійної регресії (7)

3.4 Визначення властивостей шуму і похибок оцінок параметрів

Після визначення точкових оцінок параметрів регресії розраховуємо за формулою (3) залишки регресії і результати надаємо у табличному вигляді (табл. 2).

Таблиця 2 – Залишки регресії (мг/дм³)

i	e_i	i	e_i	i	e_i
1	2	3	4	5	6
1	0,1144	11	-0,1701	21	-0,4906
2	0,1301	12	0,2812	22	-0,1623
3	0,4683	13	0,3481	23	-0,0595
4	0,2529	14	-0,3597	24	-0,4033
5	-0,165	15	0,4249	25	-0,0341
6	0,225	16	-0,3033	26	-0,3804

Продовження таблиці 2

1	2	3	4	5	6
7	-0,1749	17	0,4658	27	0,5069
8	-0,2699	18	0,1379	28	-0,1802
8	-0,2699	18	0,1379	28	-0,1802
9	0,153	19	-0,1263	29	-0,0906
10	-0,2266	20	0,3994	30	-0,3111

Також визначається стандартне відхилення залишків регресії

$$s_e = \sqrt{\frac{1}{N-2} \sum_{i=1}^N e_i^2} = 0,298 \text{ 6 мг/дм}^3.$$

Для візуальної оцінки статистичної незалежності (випадковості) залишків регресії їх розміщуємо на точковому графіку (рис. 2) разом з пунктирними лініями $e = t_s s_e = 0,611 \text{ 6}$ і $e = -t_s s_e = -0,611 \text{ 6}$, що є границями довірчого інтервалу.

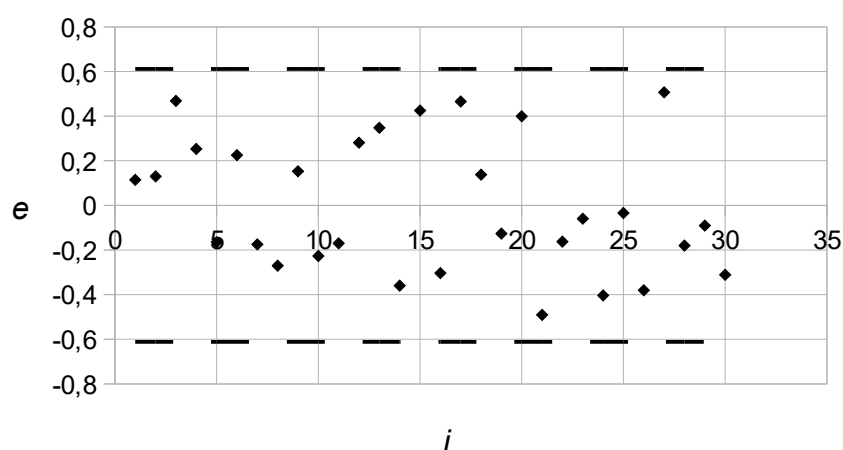


Рисунок 2 – Розподіл залишків регресії (точки, мг/дм³) в границях довірчого інтервалу

Розподіл залишків регресії на рисунку 2 виглядає досить хаотичним, неупорядкованим будь-яким чином, так що гіпотеза щодо випадковості впливу неконтрольованих факторів не відкидається. Крім того, відсутність будь-яких відхилень точок за межі довірчого інтервалу свідчить про однорідність цього впливу.

Далі розраховуємо стандартні похибки визначення параметрів регресії за формулами

$$\overline{C_{cm}^2} = \frac{1}{N} \sum_{i=1}^N C_{cm,i}^2 = 97\,053 \text{ мг}^2/\text{дм}^6;$$

$$m(\theta_0) = s_e \cdot \frac{\sqrt{\overline{C_{cm}^2}}}{s_{cm} \sqrt{N-1}} = 0,339\,8 \text{ мг}/\text{дм}^3;$$

$$m(\theta_1) = s_e \cdot \frac{1}{s_{cm} \sqrt{N-1}} = 0,001 \text{ л}.$$

Порівнюємо величини критеріїв достовірності параметрів регресії

$$\hat{\theta}_0 / m(\theta_0) = 106,0; \hat{\theta}_1 / m(\theta_1) = 9,40$$

з табличною величиною критерію Стюдента $t_s = 2,048$ з $\nu = N - 2 = 28$ ступенями вільності і рівнем значущості $\alpha = 0,05$ (дод. Б). Наявні перевищення визначають статистичну значущість обох параметрів регресії. Відповідні довірчі інтервали для цих параметрів регресії наводимо, наприклад, у такому вигляді (7, 8):

$$35,3 \text{ мг}/\text{дм}^3 \leq \theta_0 \leq 36,7 \text{ мг}/\text{дм}^3;$$

$$0,008 \leq \theta_1 \leq 0,0125,$$

Середню похибка апроксимації даних вимірювань рівнянням регресії (у відсотках) визначаємо за формулою

$$A = \frac{1}{N} \sum_{i=1}^N \left| \frac{e_i}{C_{kc,i}} \right| \cdot 100 \% = 0,67 \%;$$

Цими висновками завершуємо регресійний аналіз даних вимірювань, тобто – розробку регресійної моделі впливу скиду стічної води.

3.5 Прогнози за лінійною парною регресією

Прогноз середньої (за даними вимірювань) концентрації $\overline{C}_{kc,n}$ домішки в КС, якщо її концентрація в стічній воді дорівнює $C_{cm,n} = 180 \text{ мг}/\text{дм}^3$, надається довірчим інтервалом

$$\overline{C}_{kc,n} = \hat{\theta}_0 + \hat{\theta}_1 C_{cm,n} \pm t_s s_e \sqrt{\frac{1}{N} + \frac{(C_{cm,n} - \overline{C}_{cm})^2}{(N-1)s_{cm}^2}} =$$

$$= 36,02 + 0,010\,25 \cdot 180 \pm 2,048\,4 \cdot 0,298\,6 \cdot \sqrt{\frac{1}{30} + \frac{(180 - 307,5)^2}{29 \cdot 50,83^2}} =$$

$$= 37,86 \pm 0,306 \text{ мг/дм}^3.$$

Прогноз результату вимірювання концентрації $C_{кс,р}$ в контрольному створі, коли її концентрація в стічній воді дорівнює $C_{см,р}$, надається довірчим інтервалом

$$C_{кс,п} = \hat{\theta}_0 + \hat{\theta}_1 C_{см,п} \pm t_s s_e \sqrt{1 + \frac{1}{N} + \frac{(C_{см,п} - \bar{C}_{см})^2}{(N-1)s_{см}^2}} =$$

$$= 36,02 + 0,010\,25 \cdot 180 \pm 2,048\,4 \cdot 0,298\,6 \cdot \sqrt{1 + \frac{1}{30} + \frac{(180 - 307,5)^2}{29 \cdot 50,83^2}} =$$

$$= 37,86 \pm 0,684 \text{ мг/дм}^3.$$

3.6 Висновки

За результатами вибіркового обстеження у мов скиду стічної води підприємства у річку, які складаються з даних серійної вибірки парних вимірювань (розміром 30 пар) концентрації домішки речовини у стічній воді на випуску і в контрольному створі на річці, проведено лінійний парний регресійний аналіз, що показав:

- статистична залежність між концентрацією домішки речовини в контрольному створі на річці і її концентрацією в стічній воді на випуску існує і підтверджує вплив скиду стічної води підприємства на якість річкової води;

- виявлено статистичну значущість обох параметрів лінійної парної регресії, що разом з високою величиною коефіцієнта кореляції свідчить про високу щільність встановленої статистичної залежності;

- аналіз величин залишків регресії не виявив будь-яких закономірностей їх розподілу, що свідчить про випадковість і однорідність впливу випадкових факторів на результати вимірювань;

- рівняння регресії забезпечує досить високу точність апроксимації даних вимірювань: середня відносна похибка апроксимації не перевищує 1%;

– виконані досить точні прогнози концентрації домішки в контрольному створі за назначеною її концентрацією в стічній воді на випуску, що характеризуються малою шириною довірчих інтервалів.

4 ВИМОГИ ДО ВИКОНАННЯ РОБОТИ

Робота має бути виконана самостійно за варіантом вихідних даних, що надається викладачем. Пояснювальна записка має бути викладена українською або російською мовами.

Розрахунки виконуються за допомогою наукового калькулятора або в середовищі Excel[®], Mathcad[®].

Розрахунки за допомогою наукового калькулятора виконуються і наводяться у тексті записки з точністю до четвертої значущої цифри. Розрахунки в середовищі Excel[®] або Mathcad[®] виконуються з точністю, що дозволяє ЕОМ, але наводяться у тексті записки також з точністю до четвертої значущої цифри.

Якщо розрахунки виконувалися в середовищі Excel[®], то у додатку до пояснювальної записки наводиться роздруківка аркушів Excel[®] або Mathcad[®] з програмою.

Графіки, що створені у середовищі Excel[®], наводяться у вигляді, що представлений в розділі 3, але, якщо графіки створюються вручну, то кожний графік за розміром має бути розташований на окремому листі записки формату А4.

Робота оцінюється за такими складовими:

- самостійність у виконанні роботи;
- точність (відносні похибки) розрахунків;
- інформативність та естетика графічного відображення даних і результатів розрахунків;
- у текстовому викладені результатів курсової роботи обов'язкове додержання вимог методичних рекомендацій видання «Загальні вимоги до

підготовки, оформлення та видання наукової і навчально-методичної літератури» (методичні рекомендації / Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова; уклад. О. В. Михаленко. – 2-ге вид., зі змін. та доп. – Харків : ХНУМГ ім. О. М. Бекетова, 2017. – 66 с.).

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Сенько П. С. Теорія ймовірностей та математична статистика: монографія / П. С. Сенько; [2-ге вид., зі змін. і доп.]. – Київ : Знання, 2007. – 446 с.
2. Карташов М. В. Імовірність, процеси, статистика: монографія / М. В. Карташов. – Київ : ВПЦ Київський університет, 2007.
3. Рао С. Р. Линейные статистические методы и их применения: пер. с англ. / С. Р. Рао. – М.: Наука, 1967.

ДОДАТОК А

Критерій Фішера

Таблиця А.1 – Значення F -критерія Фішера при рівні значущості $\alpha = 0,05$

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	∞
1	2	3	4	5	6	7	8	9	10	11
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14

Продовження таблиці А.1

1	2	3	4	5	6	7	8	9	10	11
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

ДОДАТОК Б
Критерій Стюдента

Таблиця Б.1 – Значення t_s -критерія Стюдента при рівні значущості 0,10; 0,05; 0,01

Число ступенів вільності (ν)	Рівень значущості (α)			Число ступенів вільності (ν)	Рівень значущості (α)		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Виробничо-практичне видання

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
з дисципліни

«Моделювання та прогнозування стану навколишнього середовища»

до виконання контрольної роботи

**«РОЗРОБКА РЕГРЕСІЙНОЇ МОДЕЛІ ВПЛИВУ СКИДУ СТИЧНИХ ВОД
ПІДПРИЄМСТВА НА ЯКІСТЬ РІЧКОВОЇ ВОДИ
В КОНТРОЛЬНОМУ СТВОРІ»**

*(для студентів 6 курсу заочної форми навчання спеціальності
263 – Цивільна безпека)*

Укладач **БАРАННІК** Валерій Олександрович

Відповідальний за випуск *Д. В. Дядін*

За авторською редакцією

Комп'ютерне верстання *В. О. Бараннік*

План 2018, поз. 122 М

Підп. до друку 05.12.2019. Формат 60 × 84/16
Друк на ризографі. Ум. друк. арк.1,4.
Тираж 50 пр. Зам. № .

Видавець і виготовлювач:
Харківський національний університет
міського господарства імені О. М. Бекетова,
вул. Маршала Бажанова, 17, Харків, 61002.
Електронна адреса: rectorat@kname.edu.ua
Свідоцтво суб'єкта видавничої справи:
ДК № 5328 від 11.04.2017.